第21卷 第7期
2004年12月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 21 No. 7
Dec. 2004

文章编号:1005-3085(2004)07-0029-07

MS网点的合理布局

张仁丽, 李捷飞, 邱 霆
指导教师: 数模组
(浙江理工大学 310018)

编者按: 本文作者提出消费人流量的概念，并用聚类分析的方法对相应的消费金额进行讨论，由此得到20个商区内MS网点的分布方案。这种通过聚类分析来对数据做出处理和分类的方法具有其特点，值得推荐。

摘 要: 奥运会临时超市网点设计问题是一个离散的最优化设计问题，首先对所给的数据用统计学中的方法进行统计，得出观众在出行、用餐和购物等方面的规律。

在各类人流量的计算时先将各商区场地编号进行转换、再引入商区人流量特征矩阵，使无序的人流量分布在矩阵中得以有规律的表现，算出20个商区的人流量和消费流量的分布。由于各商区消费流量和消费额的相关系数ρ为0.997，认为两者线性相关，所以先对商区消费金额进行单变量聚类，初步把商区分成四类。再用人流量的指标对分类结果适当调整，运用整数规划模型得到20个商区内MS网点分布的具体方案，同类商区中MS网点分布基本上是相同的。

关键词: 消费人流量; 聚类分析

分类号: AMS(2000) 62H30　　**中图分类号**: O242.1　　**文献标识码**: A

经过分析认为影响超市选址的主要因素是商圈内的人流量及购物欲望，而在本论文中认为每个MS销售的商品一样，所以不可以简单的用人流量作为衡量商圈的消费能力，而是选择用游客的消费人流量作为其中的一个标准。一个商圈的消费人流量表示在该商圈消费的数量; 首先对所给数据进行初步分析，按照出行、餐饮和购物将数据分为3部分，可得到各自的百分比如下:

表1 各站点人数统计

	公交（南北）	公交（东西）	出租	私车	地铁（东）	地铁（西）
总人数	1774	1828	2110	958	2006	2024
百分比	16.735	17.245	18.962	9.037	18.924	19.094

表2 餐饮人数统计

	中餐	西餐	商场
总人数	2382	5567	2651
百分比	22.4717	52.5189	25.00943

表3 各消费档次人数统计

	消费1	消费2	消费3	消费4	消费5	消费6
总人数	2060	2629	4668	983	157	103
百分比	19.43396	24.80189	44.03774	9.273585	1.481132	0.971698

通过对收集的数据进行统计，得出以下几点规律：年龄档为3和4的游客偏好中餐，而年龄档为2的旅客偏好西餐；年龄与出行方式的关系不显著，因为选择使用何种交通工具主要取决

于他们使用该交通工具的方便程度；年龄对消费水平的影响比较明显，年龄档为2、3的游客中高消费人数较多；男性倾向于乘坐公共交通工具，而女性则倾向于乘坐出租车、私家车；在高消费游客中，女性明显多于男性。

把A、B、C场馆的看台编号从新编排一下：若从上面路口进入场馆，则把与路口最近的看台编为一号，顺时针依次编为2、3…K号；若从下面路口进入场馆，则把与路口最近的看台编为一号，顺时针依次编为2、3…K号。设集合$S=\{1,2,3...k\}$表示场馆的k个看台，则由置换群的理论得经过从新编号确定的置换关系

$$f_i=\begin{pmatrix} 1 & 2 & 3 & \dots & k \\ a_1 & a_2 & a_3 & \dots & a_k \end{pmatrix}$$

其中$\begin{pmatrix} a_1 & a_2 & \dots & a_k \end{pmatrix}$为1 ~ k的一个排列，$i=A,B,C$ 按照转换后做出的各个商场的消费人流量逆转换回原来各个商场的人流量的转换关系为f_i'。在一个运动场中，消费人流量由经看台A_1向与它对面的看台$A_{k/2+1}$流，则当i小于$k/2+1$时，人流必经A_i向A_{i+1}流，当i大于$\frac{k}{2}+1$时，人流经A_{k-1-i}向A_{k-i}流，则可以证明在下图中B区域与C区域对应的看台上的消费人流量是相等的（如图1），都是等概率分配的。

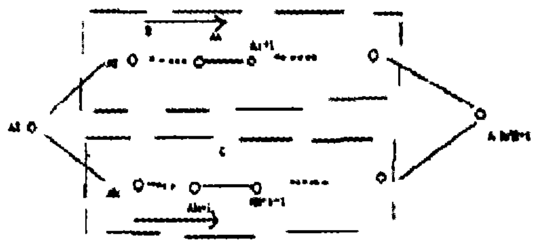

图 1: 消费人流示意

记

$$Q=\begin{bmatrix}
1 & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{i} & \frac{1}{i+1} & \dots & \frac{2}{\frac{k}{2}+1} & \frac{2}{k} & \dots & \frac{1}{2} \\
0 & \frac{1}{2} & \frac{1}{3} & \dots & \frac{1}{i} & \frac{1}{i+1} & \dots & \frac{1}{\frac{k}{2}+1} & 0 & \dots & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & \dots & 0 & 0 & \dots & \frac{2}{\frac{k}{2}+1} & 0 & \dots & 0 \\
0 & 0 & 0 & \dots & 0 & 0 & \dots & \frac{1}{\frac{k}{2}+1} & \frac{2}{k} & \dots & 0 \\
0 & 0 & 0 & \dots & 0 & 0 & \dots & \frac{1}{\frac{k}{2}+1} & \frac{2}{k} & \dots & 0 \\
0 & 0 & 0 & \dots & 0 & 0 & \dots & \frac{1}{\frac{k}{2}+1} & \frac{2}{k} & \dots & 0 \\
0 & 0 & 0 & \dots & 0 & 0 & \dots & \frac{1}{\frac{k}{2}+1} & \frac{2}{k} & \dots & \frac{1}{2}
\end{bmatrix}, \qquad
G=\begin{bmatrix}
1 & 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 & 1 \\
0 & 1 & 1 & \dots & 1 & \frac{1}{2} & 0 & \dots & 0 & 0 \\
0 & 0 & 1 & \dots & 1 & \frac{1}{2} & 0 & \dots & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\
0 & 0 & 0 & \dots & 0 & \frac{1}{2} & 1 & \dots & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & \dots & 0 & \frac{1}{2} & 0 & \dots & 0 & 0
\end{bmatrix}$$

Q是一k行k列的矩阵，Q_{ij}代表在一个场馆中第j商区对第i商区销费人流量的贡献的比例，例如编号为i的商区对编号为1的商区销售人流量的贡献率为$\frac{1}{i}$，它后面的则为$\frac{1}{i+1}$。G是一k行k列的矩阵，G_{ij}代表j商区对i商区总人流量的贡献的比例。

已经得到选择各种出行工具和各种餐馆的概率P_{ci}和P_{fi}；假设W表示总游客的人数，W_i代表第i车站的人流量其中$i=1,2,3...6$代表公交（东西）、公交（南北）、地铁（东）、地铁（西）、私车和出租）；T_i为第i种餐馆的人流量（其中i=1、2、3代表中餐、西餐和商场）；则

$$W_i=P_{ci}\times W, \qquad T_i=P_{fi}\times W$$

假设在每个车站下车的人到达三个场馆的可能性是相等的，即可以认为每个车站去3个场馆的人的概率相同，假设各车站去每个场馆的人数与场馆的容量成正比，均为$C:B:A=2:3:5$，用$M_j(j=1,2,3)$来表示各场馆容量的比例，其中$M_j=\begin{cases}0.5, & j=1;\\ 0.3, & j=2;\\ 0.2, & j=3.\end{cases}$ 而到达场馆的人是均匀散布在场馆的任意一个商区所对应的看台，认为到达场馆的各类游客是平均分布在场馆的每个商区。则从第i车站到第j场馆的k号看台的人数为

$$W_{ijk}=W_i\times\frac{1}{K_j}\times M_j$$

第i餐馆到第j场馆的第k号看台的人数为

$$T_{ijk}=T_i\times\frac{1}{K_j}\times M_j$$

由于i车站和i餐馆引起j场馆的k看台对应的出行消费流量和餐饮消费流量

$$S_{ijk}=G\times W_{ijk},\qquad R_{ijk}=G\times T_{ijk}$$

现在得到的是经过转换后的第k看台的人数S_{ijk}和R_{ijk}，对此我们对它逆转换f'_{c1}换成原来的看台旁商区的消费人流量；那么第j场馆的第k看台对应商区的总出行消费流量和总餐饮消费流量

$$S_{ik}=\sum_{i=1}^{6}S_{ijk},\qquad R_{ik}=\sum_{i=1}^{6}R_{ijk}$$

所以第j场馆的第k看台上总消费流量为

$$(S_{jk}+R_{jk})\quad j=1,2,3$$

代入数据，计算出各场馆的总人流量分布：

表4 C区总人流分布

游泳中心	c1	c2	c3	c4
出行流量（人）	14998	20707	14998	29285
餐饮流量（人）	14999	9999	14999	39996
总流量（人）	29997	30706	29997	69281
百分比（%）	2.586267	2.647395	2.586267	5.973236

表5 B区总人流分布

国家体育馆	b1	b2	b3	b4	b5	b6
出行流量（人）	20523	19471	32365	19471	20523	37623
餐饮流量（人）	22751	17246	21234	17246	22751	48759
总流量（人）	43274	36717	53599	36717	43274	86382
百分比（%）	3.730977	3.165649	4.621173	3.165649	3.730977	7.447642

表6 A区总人流分布

国家体育场	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10
出行流量（人）	50259	28418	29469	30521	31573	59723	31573	30521	29469	28418
餐饮流量（人）	30222	21740	27245	32750	38255	79767	38255	32750	27245	21740
总流量（人）	80481	50158	56714	63271	69828	139490	69828	63271	56714	50158
百分比（%）	6.9388	4.3244	4.8897	5.4550	6.0203	12.026	6.0203	5.4550	4.8897	4.3244

于是由于i车站和i餐馆引起j场馆的第k看台对应的出行人流量和餐饮人流量为

$$Z_{ijk} = Q \times W_{ijk}, \qquad H_{ijk} = Q \times T_{ijk}$$

现在得到的是经过转换后的第k看台的人数Z_{ijk} 和H_{ijk}，对此我们对它逆转换f_i' 换成原来的看台旁商区的消费人流量Z_{ijk}和H_{ijk}；那么第j场馆的第k看台对应商区的出行消费流量和餐饮消费流量为

$$Z_{jk} = \sum_{i=1}^{6} Z_{ijk}, \qquad H_{jk} = \sum_{i=1}^{3} H_{ijk}$$

所以第j场馆的第k看台上消费流量为

$$Z_{ijk} + H_{ijk}, j = 1, 2, 3.$$

代入数据，计算出各场馆的消费人流量分布：

表7 C区消费人流分布

游泳中心	c1	c2	c3	c4
出行消费流量（人）	6666	10472	6666	16190
餐饮消费流量（人）	6666	3333	6666	23331
总消费流量（人）	13332	13805	13332	39521
百分比%	3.333333	3.451595	3.333333	9.881238

表8 B区消费人流分布

国家体育馆	b1	b2	b3	b4	b5	b6
出行消费流量（人）	7346	6820	14429	6820	7346	17233
餐饮消费流量（人）	8459	5707	8492	5707	8459	23172
总消费流量（人）	15805	12527	22921	12527	15805	40405
百分比（%）	3.951645	3.132063	5.730823	3.132063	3.951645	10.10226

表9 A区消费人流分布

国家体育场	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10
出行消费流量（人）	17621	7679	6824	7175	8819	21372	8819	7175	6824	7679
餐饮消费流量（人）	9681	5268	6082	7917	11231	29315	11231	7917	6082	5268
总消费流量（人）	27302	12947	12906	15092	20050	50687	20050	15092	12906	12947
百分比（%）	6.8261	3.2370	3.2268	3.7733	5.0130	12.673	5.0130	3.7733	3.2268	3.2370

MS的设计方案必须满足在地点、大小类型和总量方面满足三个基本要求：购物需求、分布基本均衡和商业上的赢利。基于以上的要求，为了设计出一个合理的MS网点分布方案，选择消费人流量和商区日消费额这两个指标作为评判的依据。将20个商区的日消费额看成一个20维的向量$X=(X_{a1},X_{a2},X_{a3},X_{a4},X_{a5},X_{a6},X_{a7},X_{a8},X_{a9},X_{a10},X_{b1},X_{b2},X_{b3},X_{b4},X_{b5},X_{b6},X_{c1},X_{c2},X_{c3},X_{c4})$各商区$X_i$的日消费额的期望

$$E(x)=\sum_{i=6}^{n}x_i\times p(x_i)$$

由此我们把每一消费档次的均值（期望）记作m_i，由表3知道各个消费档次消费的概率，将其记作$l_i,(i=1,2,..,6)$，则由期望的理论得每个商区的日消费金额

$$X_j=\sum_{i=1}^{6}(m_i\times l_i)\times(S_{jk}+R_{jk}),(i=a1,a2,...,a10,b1,b2,...,b6,c1,...,c4);$$

代入数据得到如下金额（单位：元）：

表10 游泳中心各商区消费额

游泳中心	c1	c2	c3	c4
消费额	2677799	2772803	2677799	7937990

表11 国家体育馆各商区消费额

国家体育馆	b1	b2	b3	b4	b5	b6
消费额	3174513	2516111	4603797	2516111	3174513	8115546

表12 国家体育场各商区消费额

国家体育场	a1	a2	a3	a4	a5
消费额	5483743	2600470	2592235	3031304	4027143
国家体育场	a6	a7	a8	a9	a10
消费额	10180737	4027143	3031304	2592235	2600470

同样把20个商区消费人流量看成是一20维向量Y，由上面我们得到

$$Y_i=(Z_{jk}+H_{jk})$$

于是

$$A = X^T Y$$

对已经得到了每个场馆各个商区的日消费人流量Y以及日消费额X，按照他们在性质上的亲密程度进行划分是进行聚类的主要依据，通常对样本间的亲密程度的描述有一，把每一个样本看成是二维空间上的一点，在点与点之间定义某种距离；二，是用相似系数来描述样本间的关系。相似系数：

$$A'_{ij} = A_{ij} / \sqrt{\sum_{1}^{2} \times (A_{kj}^2)}, \quad i = 1, 2, ..., 20.\ j = 1, 2$$

则样品与样品的相似系数

$$\rho_{ij} = \sum_{k=1}^{20} A'_{kj} A'_{kj} \quad 其 i, j = 1, 2, ..., 20, \quad 且 i \neq j$$

定义样品间两点间的距离通常有，欧式距离、标准欧式距离、马式距离等，这里选择欧式距离。首先计算出X与Y相关系数$\rho \approx 1$，此时随机变量X与Y基本上线性相关，因此我们可以只考虑对日消费额X进行单变量的聚类，采用重心法对MS进行分类，得到分类冰柱图如图2：

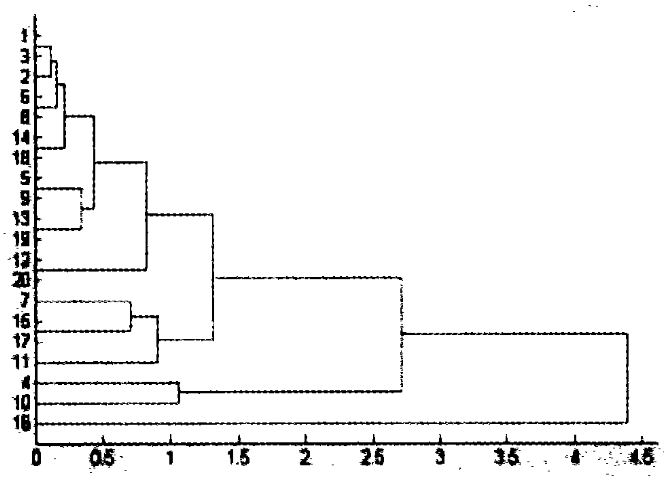

图 2：分类冰柱图

按图2找到最优为分四类，在第i类中各个商业区中超市的分布是相似的，具体的分类结果，如下：

表13　MS分类表

	第一类	第二类	第三类	第四类
分类结果	1、2、3、5、6、8、9、12、13、14、18、19、20	7、11、15、17	4、10	16

设$N_i = \begin{cases} 13, & i=1; \\ 4, & i=2; \\ 2, & i=3; \\ 1, & i=4. \end{cases}$ 表示第i类中商区的个数，则第i类的期望销售金额

$$\bar{X}_i = \sum_{j} X_j / N_i, \quad 其 j = 1, 2, ..., N_i$$

第i类的期望人流量

$$\bar{Y}_i = \sum_j Y_j/N_i, \quad 其j = 1, 2, ..., N_i$$

在这个问题中，我们认为这是一个整数线性规划模型，综合考虑商场在营业额及满足消费人流量上的目标，假设A_s表示小型超市的销售额；B_s表示小型超市的日人流量；A_l表示大型超市的销售额；B_l表示大型超市的日人流量；C_i、D_i表示第i类中大、小超市的个数；表示大型超市的销售额与小型超市的销售额之间的比例；则问题归纳为求满足方程目标函数

$$f = \min(C_i/D_i), \qquad \text{S.t.} \quad F_i \times X_i \geq \begin{bmatrix} \bar{X}_i \\ \bar{Y}_i \end{bmatrix}$$

本模型中，MS销售的货物可以分为5大类，所以认为$p = 5$，由实际生活了解到$A_s = 30$万，则

$$A_l = A_s \times p = 150万$$

最后得到具体的MS网点设计方案：

表14　MS网点设计方案表

	商区编号	大型（个）	小型（个）
第一类	c1、c2、c3、b1、b2、b4、b5、a2、a3、a4、a8、a9、a10	1	3
第二类	b3、a1、a5、a7	2	3
第三类	c4、b6	4	6
第四类	a6	5	9

参考文献：

[1] 叶中行等. 概率论与数理统计[M]. 科学出版社，2003年

[2] 邵学才. 离散数学[M]. 电子工业出版社，2003年

[3] 刘则毅. 科学计算技术与Matlab[M]. 科学出版社，2001年

[4] 唐守正. 多元统计分析方法[M]. 中国林业出版社，1986年

Reasonable Design of MS Site

ZHANG Ren-li, LI Jie-fei, QIU Ting

Counselor: Mathematics Modeling Group

(Zhejiang University of science, Zhejiang 310018)

Abstract: The design of a temporary supermarket network during the Olympic games is a matter of dispersed optimization. First we get the patterns of spectators' touring, dining and shopping through statistical analysis. To work out the distribution of population and consumption flow in 20 commercial areas, we set up a matrix of character after transforming the coding of commercial areas in which the disordered distribution of population flow can be seen in regularity. We find that the consumption flow and the consumption volume is linearly correlated (the correlation coefficient ρ =0.997). Then we make single variable clustering of the consumption volume of each commercial area, dividing the commercial areas into four categories that are adjusted in light of the population flow. Finally we work out the MS network project for the 20 commercial areas by establishing a round number model. It can be seen that in each category the distribution of MS networks is nearly identical.

Keywords: the distribution of population and consumption flow; cluster analysis